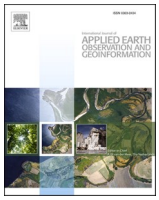
International Journal of Applied Earth Observations and Geoinformation 108 (2022) 102727

Contents lists available at ScienceDirect

International Journal of Applied Earth Observations and Geoinformation

journal homepage: www.elsevier.com/locate/jag



Estimating species-specific leaf area index and basal area using optical and SAR remote sensing data in Acadian mixed spruce-fir forests, USA

Rajeev Bhattarai [a], Parinaz Rahimzadeh-Bajgiran [a,*], Aaron Weiskittel [a], Saeid Homayouni [b], Tawanda W. Gara [a], Ryan P. Hanavan [c]

[a] School of Forest Resources, College of Natural Sciences, Forestry, and Agriculture, University of Maine, Orono, ME 04469, USA
[b] Institut National de la Recherche Scientifique, Centre Eau Terre Environment, Quebec G1K 9A9, Canada
[c] USDA Forest Service, Missoula, MT 59804, USA

ARTICLE INFO

Keywords:
Forest inventory
Red spruce
Balsam fir
Random Forest
Multi-Layer Perceptron
Remote sensing
Maine

ABSTRACT

This study combined Sentinel-1 synthetic aperture radar (SAR), Sentinel-2 multispectral, and site variable datasets to model leaf area index (LAI) and basal area per ha (BAPH) of two economically important tree species in Northeast, USA; red spruce (*Picea rubens* Sarg.; RS), and balsam fir (*Abies balsamea* (L.) Mill.; BF). We used Random Forest (RF), and Multi-Layer Perceptron (MLP) algorithms for LAI and BAPH modeling. The results showed that RF outperformed MLP by reducing the normalized root mean square error (nRMSE) by 0.01 and 0.06 for LAI and BAPH, respectively. The final variables selected for modeling of both LAI and BAPH indicated the superiority of Sentinel-2 variables over the Sentinel-1 SAR with minor contributions of site variables (mainly elevation). The red-edge spectral vegetation indices played a significant role in both LAI and BAPH estimation. We attained the lowest nRMSEs of 0.12, and 0.16 for the final LAI model of RS, and BF, respectively using Sentinel-2 and site variables. The lowest nRMSE for both RS and BF BAPH models was 0.12. As RS and BF are the primary host species for a cyclically occurring and most destructive pest of the region, eastern spruce budworm (*Choristoneura fumiferana*; SBW), these estimations will be useful to evaluate SBW dynamics in the region.

1. Introduction

Forest resources play a vital role in the long-term fulfillment of societal needs and ensuring the continued availability of ecological services; however, forest ecosystems around the world are under increasing threats from several stressors such as pests and diseases, climate change, and shifts in land use. Therefore providing comprehensive information about key forest attributes and dynamics is essential for strategic planning and sustainable forest ecosystem management. Forest stand-level inventory attributes such as leaf area index (LAI) and basal area per ha (BAPH) are primarily of interest to both the research community and land managers because they are significant indicators of forest condition, structure, and function (Fassnacht et al., 1994; Mensah et al., 2020).

LAI is considered as a determinant of forest characteristics like vegetation growth, stocking, and wood productivity, which also largely correlates with soil water storage and transpiration (Running, 1992), photosynthetic capacity (Mensah et al., 2020), the magnitude of carbon and energy fluxes (Fassnacht et al., 1994), as well as many more biological and ecological phenomena on stand, regional, and global-scales. BAPH is another central attribute of a forest stand that needs to be considered for sustainable management, as it is a measure of forest density and degree of competition (Ahmadi et al., 2020). LAI and BAPH measurement and mapping hold a long history, which have been widely completed using traditional field measurement methods (Fang et al., 2019). Direct measurement of these parameters using optical or destructive sampling in the field is neither feasible nor desirable on a regional or global scale (Fassnacht et al., 1994; Dube et al., 2019). With the advent of freely available satellite data, remote sensing techniques are gaining popularity among researchers for a rapid, accurate, and efficient estimation of forest stand attributes and their temporal evolution on a landscape level (Fang et al., 2019).

There exists a current suite of non-commercial remote sensing data to derive the LAI and BAPH in forests. Single-date optical data such as Landsat (Dube et al., 2019; Neinavaz et al., 2019), Moderate Resolution Imaging Spectroradiometer (MODIS) (Sprintsin et al., 2007; Qiao et al.,

* Corresponding author.
E-mail address: parinaz.rahimzadeh@maine.edu (P. Rahimzadeh-Bajgiran).

https://doi.org/10.1016/j.jag.2022.102727
Received 31 October 2021; Received in revised form 26 January 2022; Accepted 16 February 2022
Available online 24 February 2022

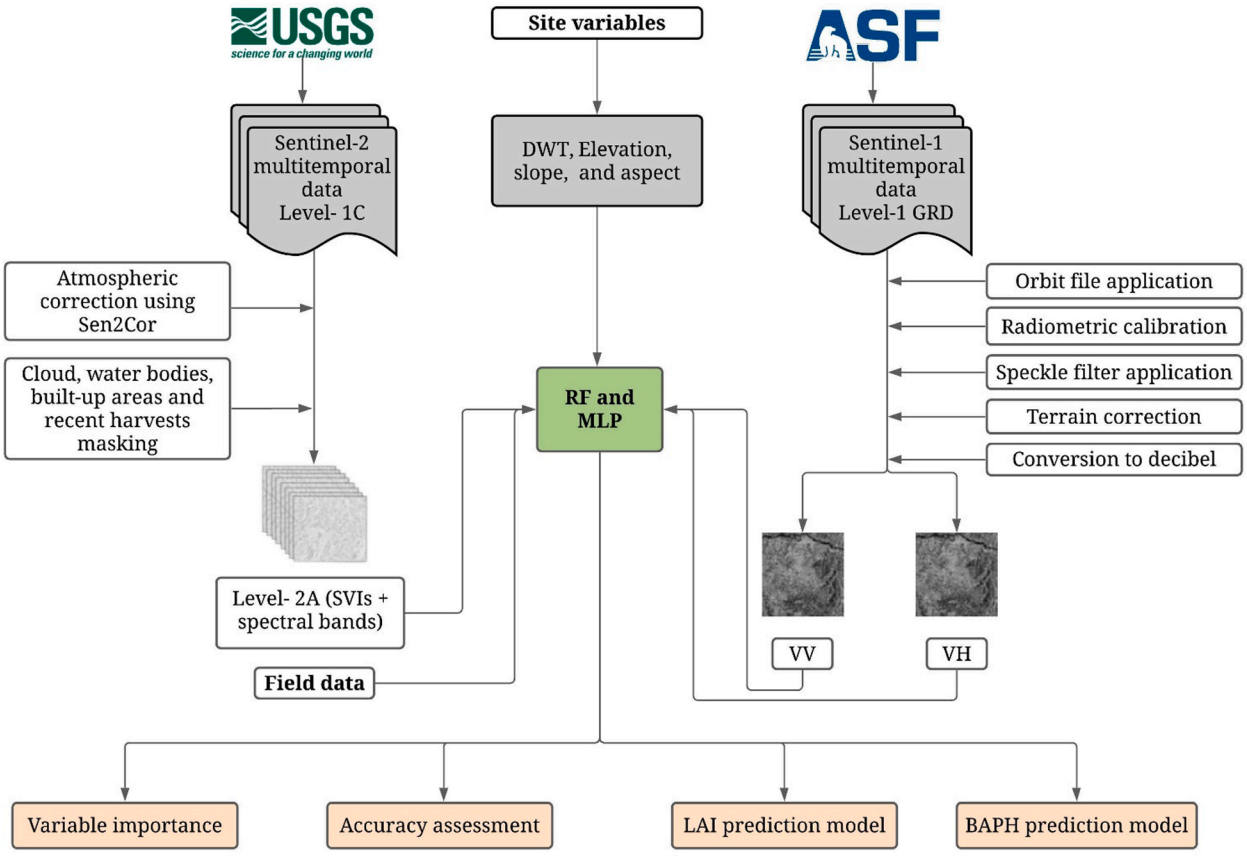


Fig. 1. General workflow of the methodology adopted to model leaf area index (LAI), and basal area per hectare (BAPH). VV and VH are the Vertical-Vertical, and Vertical-Horizontal polarization modes, respectively. RF: Random Forest; MLP: Multi-Layer Perceptron; DWT: depth to water table; SVIs: spectral vegetation indices.

2019), and Sentinel-2 (Korhonen et al., 2017; Mananze et al., 2018; Darvishzadeh et al., 2019) have been widely used for the estimation of LAI. However, a recent trend depicts the preference of Sentinel-2 data over other satellite data sources attributed to their decent and even finer resolution (spatial, spectral, and temporal) at no acquisition cost (Korhonen et al., 2017; Darvishzadeh et al., 2019; Meyer et al., 2019). In addition, the use of synthetic aperture radar (SAR) data has also been reported for LAI modeling (Manninen et al., 2005; Stankevich et al., 2017), which can be a complementary data source for the optical data. However, the combined use of SAR and optical data for modeling of LAI across diverse or mixed forest areas has not been reported. In contrast, BAPH modeling has been done primarily using either optical data such as Landsat (Wolter et al., 2008), and Sentinel-2 (Astola et al., 2019; Ahmadi et al., 2020) or SAR data (Townsend, 2002). To our knowledge only one study was reported that used a combination of the two types of data (Wolter and Townsend, 2011). Given that the optical and SAR data represent signals from different regions of the electromagnetic spectrum, and are complementary to each other, the incorporation of SAR data is expected to add structural information of tree canopy to the optical data that mainly reflect biochemical properties (Wolter and Townsend, 2011; Bhattarai et al., 2021).

While BAPH estimation is widely performed using statistical modeling utilizing its relationship with remote sensing variables, particularly spectral vegetation indices (SVIs) (Wolter et al., 2008; Ahmadi et al., 2020), the estimation of LAI is generally based on either of the following methods: 1) empirical methods that statistically relate LAI to canopy reflectance or SVIs; or 2) physical or radiative transfer models where LAI is retrieved through inversion of a spectral library created in forward mode (Delegido et al., 2011; Fang et al., 2019; Darvishzadeh et al., 2019). Empirical methods are often biome- and/or time-specific and can suffer from saturation effect (Delegido et al., 2011); but this can be addressed by including appropriate site variables and optimized band combinations for creating SVIs, in particular red-edge and three-band combination SVIs (Mananze et al., 2018). In addition, using a combination of suitable SVIs and robust machine learning algorithms, empirical models can outperform physical ones due to their statistical superiority (Cui and Kerekes, 2018; Manzane et al., 2018). Furthermore, empirical methods are relatively simple and straightforward for both interpretation and future development (Delegido et al., 2011; Fang et al., 2019).

Our study focused on the estimation of both LAI and BAPH for two commercially and ecologically important tree species in eastern North America, balsam fir (*Abies balsamea* (L.) Mill.; BF) and red spruce (*Picea rubens* Sarg.; RS), which are also threatened by a periodic outbreak of eastern spruce budworm (*Choristoneura fumiferana* Clem.; SBW) (Rahimzadeh-Bajgiran et al., 2018). The susceptibility of the forest stands to SBW is largely impacted by the landscape-scale distribution and stand-level characteristics of the specific host species (Wolter et al., 2008; Chen et al., 2021). In general, the estimation of host species abundance (BAPH) is essential for the study of SBW dynamics (Wolter et al., 2008; Wolter and Townsend, 2011). Besides, LAI is believed to be a better tool for management planning in multiage mixed-conifer forests (O'Hara and Battles, 2020). A comprehensive inventory of the primary host species can therefore guide us to identify areas with high host-species abundance. Overall, we aimed to accurately estimate the abundance of the two commercially important tree species in Acadian mixed spruce-fir forests using remote sensing techniques. Our specific objectives were:

- Modeling the LAI and BAPH of both BF and RS using the combined application of Sentinel-1 SAR and Sentinel-2 optical satellite data.

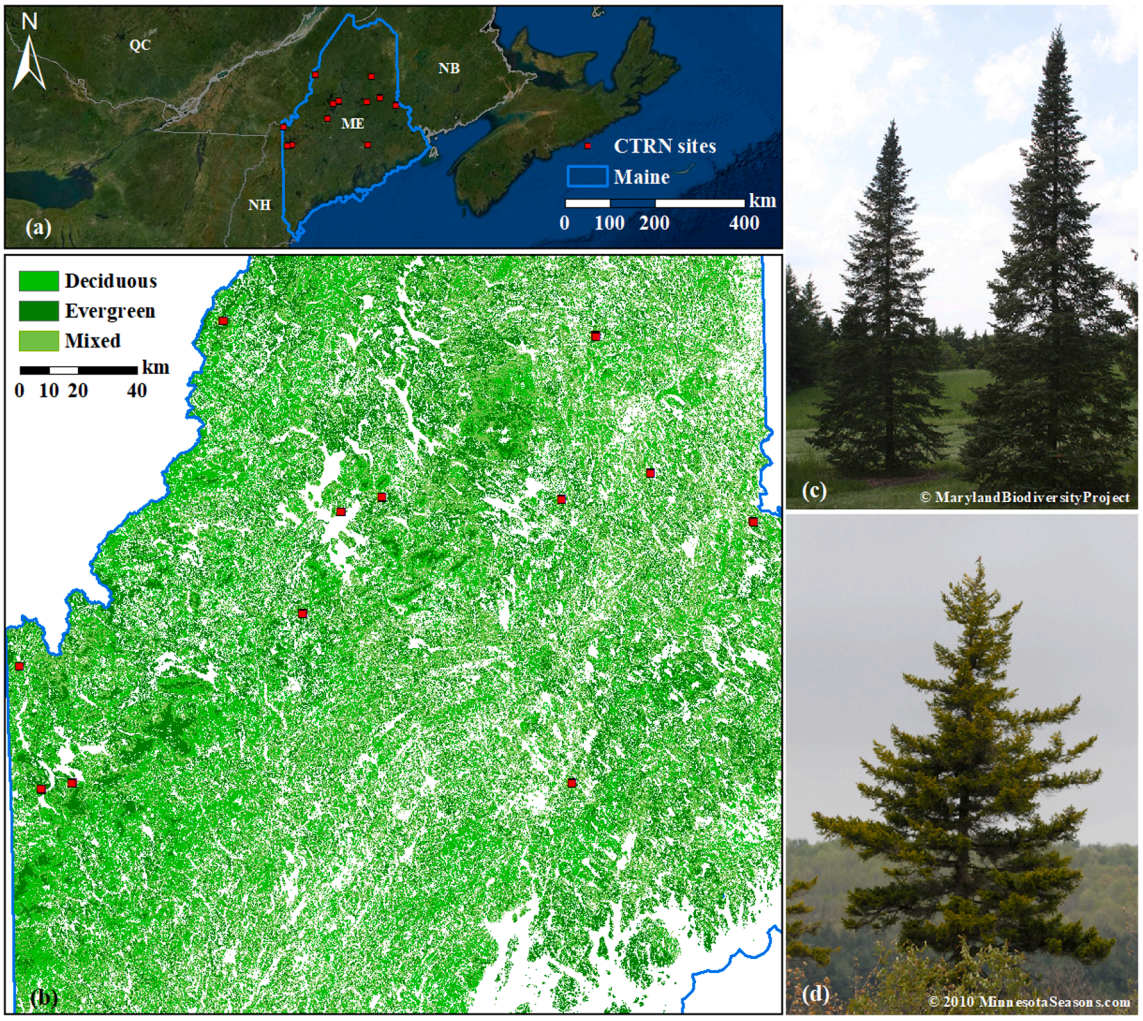


Fig. 2. Location of University of Maine Cooperative Forestry Research Unit's Commercial Thinning Research Network (CTRN) sites (a) and forest type map retrieved from the National Land Cover Database (NLCD 2016) (b) where the evergreen forest type is dominated by balsam fir (c) and red spruce (d). QC: Quebec; NB: New Brunswick; ME: Maine; NH: New Hampshire.

- Evaluating the contribution of site variables to the species-specific estimation of LAI and BAPH.
- Improving the accuracy of LAI and BAPH modeling through incorporating deep learning algorithms.
- Identifying key predictors for LAI and BAPH in diverse spruce-fir forests of Maine.

2. Materials and methods

The workflow presented in Fig. 1 summarizes the entire procedure of our study. Explanations of each step in the workflow are described in later sections.

2.1. Study area and field data

This study was conducted in the state of Maine, USA (Fig. 2). Maine has 89.1 percent of the total land area covered with forest, where BF and RS are the most abundant tree species in terms of number (36 percent of the entire stems) and volume (83.2 percent of the total tree volume), respectively (Butler, 2018).

The ground plot data were collected from 12 University of Maine Cooperative Forestry Research Unit's Commercial Thinning Research Network (CTRN) sites (Fig. 2) in northern and central Maine in summer 2018 (Kuehne et al. 2016). We used a total of 73 measurement plots (26.6×30.5 m^2) with 40 dominant BF and 33 dominant RS plots. The tree-level measurements: diameter at breast height (DBH), crown length (CL), tree height (HT) and tree leaf area (TLA) were measured in each plot (DeRose and Seymour, 2010). TLA was estimated for each tree in the plot using a species- and site-specific allometric equation (Weiskittel et al., 2009):

$$TLA = b_0 * DBH^{\left(b_1 + b_2 * CL + b_3 * \left(\frac{HT}{DBH}\right)\right)} \quad (1)$$

where b_0, b_1, b_2, and b_3 are the unique parameters of the site- and species-specific, tree-level leaf area equation for different tree species with remaining variables already defined above. Finally, LAI was derived using the following equation:

$$\text{LAI} = \text{TLA/plot area} \quad (2)$$

BAPH for each plot was estimated using the DBH. The minimum/maximum values for the LAI and BAPH of inventory plots were 0.22 m^2 m^{-2}/5.85 m^2 m^{-2} and 2.22 m^2 ha^{-1} /48.33 m^2 ha^{-1}, respectively.

2.2. Satellite data collection and preprocessing

We downloaded 21 multi-temporal Sentinel-2A and 2B Level 1C Top-Of-Atmosphere (L1C-TOA) reflectance images (https://earthexplorer.usgs.gov/) between 2017 and 2020 and synthetic aperture radar

Table 1
Sentinel-2 derived SVIs used for RS and BF LAI and BAPH modeling.

SVIs	Equation
Structure	
Atmospherically Resistant Vegetation Index (ARVI)	(B8A-2*B4 + B2)/(B8A + 2* B4 + B2)
Enhanced Vegetation Index7 (EVI7)	2.5*(B7-B4)/(1 + B7 + 6*B4-7.5*B2)
Enhanced Vegetation Index8 (EVI8)	2.5*(B8A-B4)/(1 + B8A + 6* B4-7.5*B2)
Modified Simple Ratio (MSR)	((B7/B4)-1)/sqrt((B7/B4) + 1)
Normalized Difference Vegetation Index (NDVI)	(B8A-B4)/(B8A + B4)
Soil Adjusted Vegetation Index (SAVI)	1.5*(B8A-B4)/(B8A + B4 + 0.5)
Wide Dynamic Range Vegetation Index (WDRVI)	((0.01*B7)-B4) /((0.01*B7 + B4)+((1–0.01)/(1 + 0.01)))
Physiology/stress	
Anthocyanin Reflectance Index1 (ARI1)	(1/B3)-(1/B5)
Anthocyanin Reflectance Index2 (ARI2)	(B8A/B3)-(B8A/B5)
Carotenoid Reflectance Index1 (CRI1)	(1/B2)-(1/B3)
Carotenoid Reflectance Index2 (CRI2)	(1/B2)-(1/B5)
Normalized Difference Infrared Index11 (NDII11)	(B8A-B11)/(B8A + B11)
Normalized Difference Infrared Index 12 (NDII12)	(B8A-B12)/(B8A + B12)
Plant Senescence Reflectance Index (PSRI)	(B4 − B3)/B8A
Biochemistry	
Chlorophyll Green Index (GCI)	(B8A/B3) –1
Chlorophyll Red Edge (Clre)	(B7/B5)-1
Green Atmospherically Resistant Index (GARI)	(B8A-B3-(B2-B4))/(B8A + B3-(B2-B4))
Green NDVI (GNDVI)	(B8A-B3)/(B8A + B3)
Inverted Red Edge Chlorophyll Index (IRECI)	(B7-B4)*(B6/B5)
MERIS Terrestrial Chlorophyll Index (MTCI)	(B6-B5)/(B5-B4)
Modified Chlorophyll Absorption in Reflectance Index (MCARI)	1-((0.2)*(B5-B3)/(B5-B4))
Normalized Difference Vegetation Index45 (NDVI45)	(B5-B4)/(B5 + B4)
Normalized Difference Vegetation Index65 (NDVI65)	(B6-B5)/(B6 + B5)
Red-Edge Normalized Difference Vegetation Index (NDVIRE)	(B8A-B6)/(B8A + B6)
Sentinel-2 Red Edge Position (S2REP)	705 + 35*((((B7 + B4)/2)-B5)/(B6-B5))
Transformed Chlorophyll Absorption in Reflectance Index (TCARI)	3*((B5-B4)-0.2*(B5-B3)*(B5/B4))
Triangular Vegetation Index (TVI)	0.5*(120*(B6 − B3) − 200*(B4 − B3)

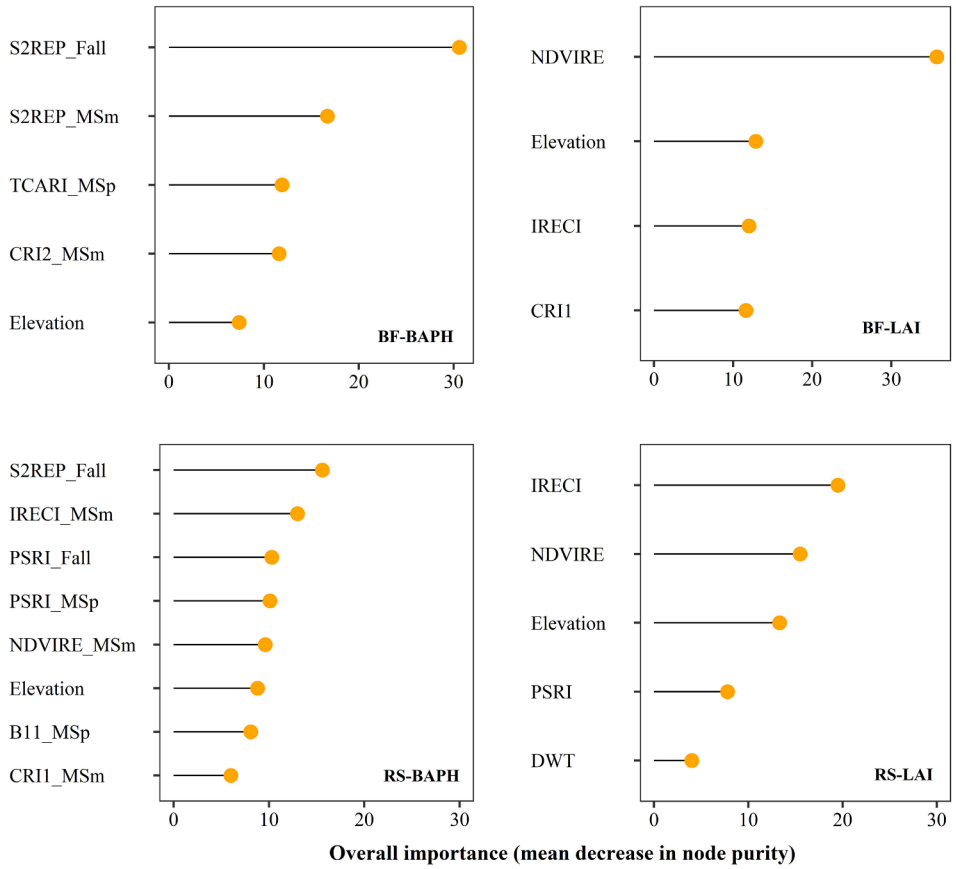


Fig. 3. Variable importance plots for four final models using Sentinel-1, Sentinel-2, and site variables. Variable abbreviations follow the pattern of, variable name_season of acquisition. MSp: mid-spring; MSm: mid-summer; Fall: fall (refer to Table 1 for the variables names and information).

Table 2
Performance of the models formulated to estimate LAI and BAPH using RF model. DWT: depth to the water table. SAR includes six variables (VV_MSp, VH_MSp, VV_MSm, VH_MSm, VV_Fall, VH_Fall) for BAPH while two variables (VV_MSm, VH_MSm) for LAI modeling.

Models	Variables used	R^2 (sd)	RMSE (sd)	nRMSE (sd)
RS-LAI	IRECI	0.45 (0.15)	0.95 (0.15)	0.18 (0.03)
	NDVIRE	0.45 (0.21)	1.00 (0.28)	0.19 (0.05)
	PSRI	0.25 (0.16)	1.10 (0.20)	0.21 (0.04)
	IRECI + Elevation	0.68 (0.18)	0.75 (0.21)	0.15 (0.04)
	IRECI + NDVIRE + PSRI	0.67 (0.17)	0.74 (0.17)	0.14 (0.03)
	IRECI + NDVIRE + PSRI + Elevation	0.74 (0.16)	0.65 (0.16)	0.13 (0.03)
	IRECI + NDVIRE + PSRI + Elevation + DWT	**0.75 (0.16)**	**0.63 (0.19)**	**0.12 (0.04)**
	IRECI + NDVIRE + PSRI + SAR	0.69 (0.17)	0.72 (0.20)	0.14 (0.04)
	SAR	0.11 (0.13)	1.00 (0.20)	0.29 (0.06)
	SAR + Elevation	0.10 (0.13)	0.93 (0.17)	0.27 (0.05)
BF-LAI	NDVIRE	0.59 (0.15)	1.31 (0.23)	0.22 (0.04)
	IRECI	0.36 (0.15)	1.66 (0.15)	0.28 (0.04)
	CRI1	0.26 (0.18)	1.84 (0.36)	0.31 (0.06)
	NDVIRE + Elevation	0.71 (0.13)	1.06 (0.23)	0.18 (0.04)
	NDVIRE + IRECI + CRI1	0.71 (0.13)	1.09 (0.21)	0.19 (0.04)
	NDVIRE + IRECI + CRI1 + Elevation	**0.78 (0.10)**	**0.95 (0.19)**	**0.16 (0.03)**
	NDVIRE + IRECI + CRI1 + Elevation + DWT	0.77 (0.09)	0.97 (0.19)	0.17 (0.03)
	NDVIRE + IRECI + CRI1 + SAR	0.72 (0.11)	1.07 (0.20)	0.18 (0.03)
	SAR	0.13 (0.13)	1.08 (0.19)	0.32 (0.06)
	SAR + Elevation	0.12 (0.14)	1.05 (0.17)	0.32 (0.05)
RS-BAPH	S2REP_Fall + IRECI_MSm + PSRI_Fall + PSRI_MSp + NDVIRE_MSm + B11_MSp + CRI1_MSm	0.79 (0.12)	5.27 (1.29)	0.12 (0.03)
	S2REP_Fall + IRECI_MSm + PSRI_Fall + PSRI_MSp + NDVIRE_MSm + B11_MSp + CRI1_MSm + Elevation	**0.80 (0.11)**	**5.18 (1.23)**	**0.12 (0.03)**
	S2REP_Fall + IRECI_MSm + PSRI_Fall + PSRI_MSp + NDVIRE_MSm + B11_MSp + CRI1_MSm + Elevation + DWT	0.80 (0.12)	5.20 (1.32)	0.12 (0.03)
	S2REP_Fall + IRECI_MSm + PSRI_Fall + PSRI_MSp + NDVIRE_MSm + B11_MSp + CRI1_MSm + SAR	0.79 (0.10)	5.47 (1.33)	0.13 (0.03)
	SAR	0.11 (0.14)	8.77 (1.87)	0.29 (0.06)
	SAR + Elevation	0.14 (0.15)	8.55 (1.84)	0.28 (0.06)
BF-BAPH	S2REP_Fall + S2REP_MSm + TCARI_MSp + CRI2_MSm	0.87 (0.06)	5.71 (1.45)	0.12 (0.03)
	S2REP_Fall + S2REP_MSm + TCARI_MSp + CRI2_MSm + Elevation	**0.87 (0.06)**	**5.69 (1.44)**	**0.12 (0.03)**
	S2REP_Fall + S2REP_MSm + TCARI_MSp + CRI2_MSm + Elevation + DWT	0.88 (0.06)	5.85 (1.27)	0.13 (0.03)
	S2REP_Fall + S2REP_MSm + TCARI_MSp + CRI2_MSm + SAR	0.84 (0.08)	6.40 (1.49)	0.14 (0.03)
	SAR	0.17 (0.15)	7.99 (1.11)	0.30 (0.04)
	SAR + Elevation	0.15 (0.14)	8.11 (1.17)	0.30 (0.04)

(SAR) Sentinel-1 images (https://asf.alaska.edu/) for 2018. Multi-temporal images for three seasons of mid-spring (May, MSp), mid-summer (June and July, MSm) and fall (September through November, Fall) were used for modeling BAPH, whereas LAI estimation was conducted using single-date imagery from the peak growing season (June-July).

We used the Sen2Cor plugin in Sentinel Application Platform (SNAP) to convert Sentinel-2 L1C-TOA products to L2A surface reflectance products and then to calculate several SVIs (Table 1) at 20 m spatial resolution (Korhonen et al., 2017; Dube et al., 2019; Bhattarai et al., 2021). Our variables include both spectral bands and SVIs, resulting in 118 Sentinel-2 variables from three dates for BAPH estimation, and 36 variables (single-date) for LAI estimation. The dual-polarized (VH: Vertical-Horizontal, and VV: Vertical-Vertical) Sentinel-1 data were preprocessed in SNAP, then resampled to 20 m spatial resolution (Bhattarai et al. 2021). In addition to single-date speckle filtering, we also incorporated multi-temporal speckle filtering in the preprocessing workflow of SAR imagery, considering the promising results from previous studies (Yuan et al., 2018). The multi-temporal speckle filtering method was expected to better suppress the speckles in SAR images without compromising the spatial resolution, taking advantage of temporal information from the images. However, in our study, multi-temporal speckle filtering did not produce better results (results not included) than the single-date speckle filtering; consequently, models were built using the data from single-date speckle filtering. We ultimately derived two sets of six (three-date; VV_MSp, VH_MSp, VV_MSm, VH_MSm, VV_Fall, VH_Fall) and two (single-date; VV_MSm, VH_MSm) variables from Sentinel-1 for BAPH and LAI modeling, respectively.

2.3. Modeling LAI and BAPH

Artificial neural networks (ANN) (Jain et al., 1996) were used and compared with the widely used machine learning algorithm, Random Forest (RF) (Breiman, 2001), for modeling LAI and BAPH estimation. We used a feedforward ANN known as Multi-Layer Perceptron (MLP), which is trained with a backpropagation technique (Jain et al., 1996). From the application point of view, MLP is becoming popular among researchers in various realms, including analyzing satellite data (Foody, 2004; Astola et al., 2019). On the other hand, RF is an ensemble of decision trees proven suitable in remote sensing data analysis due to its interpretation simplicity and usability with different data types (Astola et al., 2019; Bhattarai et al., 2021).

The VSURF R-package was used for Sentinel-2 variable reduction based on the recursive variables elimination technique (Bhattarai et al., 2021). In addition, the Pearson correlation coefficient (ranging between −1 and +1) between the variables returned after running the VSURF algorithm was evaluated, and highly correlated variables (correlation coefficient approaching −1 or +1) were eliminated. Eventually, the SAR and site variables (elevation, slope, aspect, and depth to water table (DWT) at 10 m spatial resolution) (White et al., 2013) were incorporated with the Sentinel-2 remote sensing variables to reinforce the model. MLP and RF models were run with the final sets of variables in R 3.5.1 using the Caret package, compared and the best model was used for the final prediction.

We selected the two final individual models with the best overall accuracy as well as behavior to estimate LAI and BAPH for RS and BF. Final models were validated based on the repeated k-fold cross-

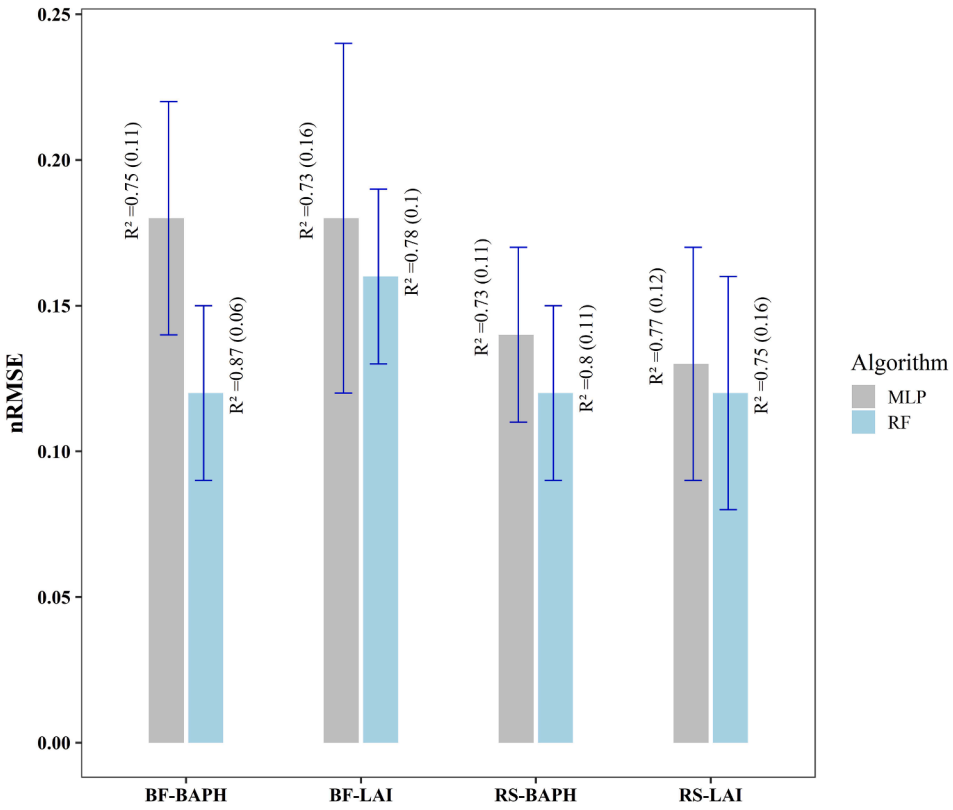


Fig. 4. LAI, and BAPH model accuracy for the RF and MLP models using the best performing variables obtained after the variable reduction procedure (Table 2 and Fig. 3). Accuracy metrics are presented with standard deviation (sd) inside the parenthesis. Error bars represent one sd.

validation technique, where five-fold cross-validation was repeated five times. External validation was carried out on the test data (30% of total data), which were never used for model training. Furthermore, the performance of the models was assessed based on the coefficient of determination (R^2), root mean squared error (RMSE), and normalized RMSE (nRMSE: normalized with range), along with their standard deviations (sd).

3. Results

3.1. Selection of Sentinel-1, Sentinel-2, and site variables for modeling

The final set of variables (including Sentinel-1, Sentinel-2, and site variables) as shown in Fig. 3 depicted the dominance of Sentinel-2 variables for the RS, and BF LAI and BAPH modeling. Most of the Sentinel-2 variables selected for the prediction were red-edge variables and represented the category of SVIs expressing physiology/stress (CRI2, PSRI, CRI1) and leaf chemistry (S2REP, TCARI, IRECI, and NDVIRE). Specifically, S2REP and IRECI were prominent for modeling BAPH, while NDVIRE and IRECI were significant for LAI. In addition, the variables selected for BAPH estimation mainly consisted of mid-summer as compared to mid-spring and fall variables.

The incorporation of Sentinel-1 variables did not benefit any of the final models, although these variables demonstrated some promise on integration with Sentinel-2 variables during the intermediate formulations of LAI models (Table 2). However, the site variables were significant for all models. Elevation was among the best variables for modeling LAI and BAPH of both species, while DWT was important for modeling LAI for RS only. Slope and aspect were not influential for any of the models.

3.2. Model performance and validation

Several combinations of Sentinel-1, Sentinel-2, and site variables were tested for LAI and BAPH modeling using the RF algorithm, and the outcomes are presented in Table 2. In this table, the rows in bold indicate our best model combinations based on Fig. 3. However, to be comparable with available literature, we also evaluated single best SVI and SAR variables for model construction.

The RF models developed for LAI estimation attained the lowest nRMSE of 0.14 and 0.18 for RS and BF, respectively, with Sentinel-2 multi-variable combination. The superiority of the multi-variable model was observed over models based on a single SVI for LAI estimation in our study (Table 2). Further addition of SAR variables marginally decreased the estimation error (0.72 $m^2\ m^{-2}$ for RS and 1.07 $m^2\ m^{-2}$ for BF) with nRMSE almost the same; however, the combination of Sentinel-2 and site variables generated better results. The combination of Sentinel-2 and site variables (Elevation and DWT) attained an overall lowest LAI estimation nRMSE of 0.12 for RS, whereas Sentinel-2 variables plus elevation yielded the lowest estimation nRMSE of 0.16 for BF.

BAPH for both BF and RS was estimated with nRMSE of 0.12, using Sentinel-2 variables alone. The addition of Sentinel-1 variables did not improve the accuracy further. Nevertheless, incorporating site variables (elevation) decreased the RMSE slightly to 5.69 $m^2\ ha^{-1}$ for BF-BAPH and 5.18 $m^2\ ha^{-1}$ for RS-BAPH with the nRMSE unchanged at 0.12. Besides elevation, none of the site variables (slope, aspect, or DWT) helped to improve the model.

Sentinel-1 variables alone performed poorer than expected with RF (Table 2). However, the addition of site variables, especially elevation, improved the model accuracy marginally for all models (RS-LAI, BF-LAI, and RS-BAPH) except BF-BAPH, where the model error increased after the addition of elevation. After incorporating elevation, the lowest nRMSEs achieved with Sentinel-1 variables were 0.27, 0.32, 0.28, and 0.30 for RS-LAI, BF-LAI, RS-BAPH, and BF-BAPH models, respectively.

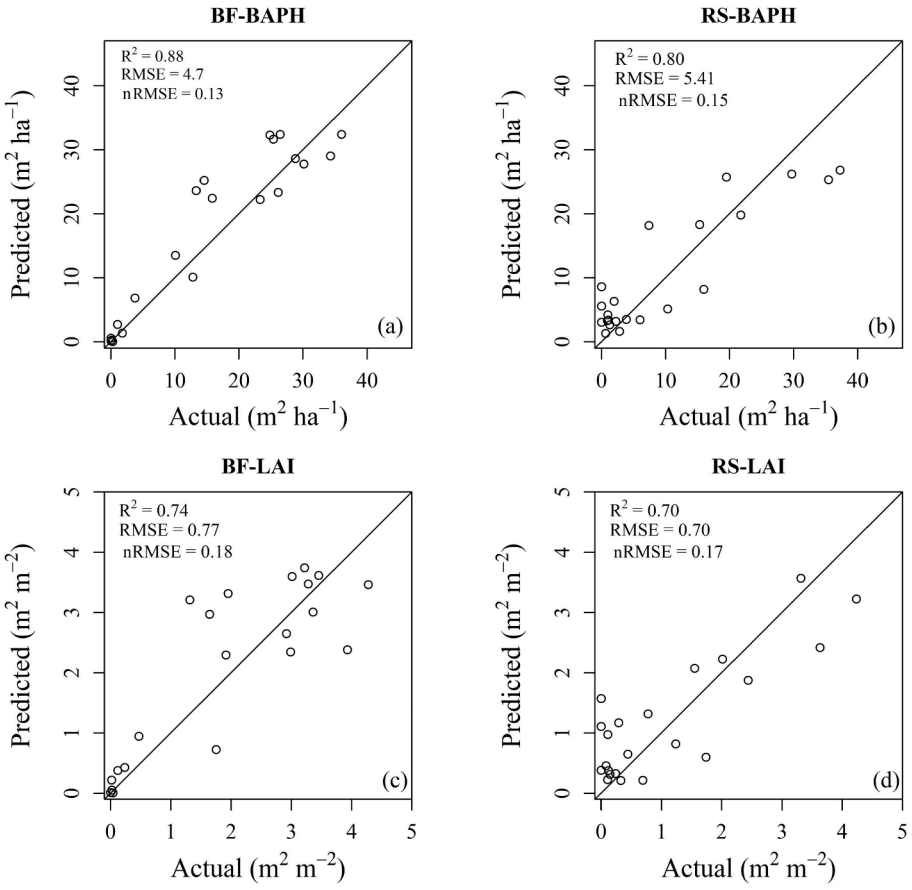


Fig. 5. Actual vs. predicted values of LAI and BAPH for four best RF regression models (Table 2). The diagonal line running across each subplot is a 1:1 relationship between the actual and predicted values.

Under further evaluation of the performance of final variables obtained after variable reduction using MLP regression (Fig. 3), none of the formulated models could yield better accuracy than their RF regression counterparts (Fig. 4). A single hidden layer returned the best results for all four MLP models. RS-LAI and BF-LAI attained the best result with six and five neurons, respectively, while RS-BAPH and BF-BAPH required seven individual neurons in the hidden layer. The nRMSEs with the MLP algorithm for RS-LAI, BF-LAI, RS-BAPH, and BF-BAPH models were 0.13 (0.04), 0.18 (0.06), 0.14 (0.03), and 0.18 (0.04), respectively.

The four best RF models presented in bold in Table 2 were selected to predict the independent test dataset. The model performance on the testing dataset (Fig. 5) confirms the stability of the models we created using our training data. We obtained a robust nRMSE as low as 0.13 and 0.15 for BF-BAPH, and RS-BAPH whereas these values were 0.18 and 0.17 for BF-LAI, and RS-LAI, respectively. Fig. 5 demonstrates the relationship between the predicted and actual LAI and BAPH values, which highlights our models' capacity on a relatively independent dataset.

4. Discussion

Our study attempted to construct models to estimate the species-specific LAI and BAPH of two economically important tree species in Maine, namely, red spruce (RS) and balsam fir (BF). Previous studies on the estimation of LAI (Korhonen et al., 2017; Neinavaz et al., 2019) and BAPH (Ahmadi et al., 2020) of forest cover were generalized and not species-specific. The species-specific models to estimate LAI and BAPH of our analysis are more useful in particular to determine the susceptibility of forests to host-specific defoliators like SBW and to evaluate the basal area of the hosts under risk, respectively. In addition, we evaluated the combined use of SAR and optical data for LAI and BAPH modeling, while also including several site variables in their estimation.

We used combined Sentinel-2 multispectral, Sentinel-1 SAR, and site data with widely used machine learning algorithms (RF and MLP). Our results demonstrate that incorporating Sentinel-1 SAR variables did not improve models based on Sentinel-2 data using either single-date or a multi-temporal speckle filtering procedure. Our results in terms of the performance of Sentinel-1 SAR data for LAI estimation are in contrast to the findings from Manninen et al. (2005). They reported the superiority of radar (Environmental Satellite (ENVISAT) Advanced Synthetic Aperture Radar (ASAR) C-band) data over optical (SPOT high-resolution visible and infrared (HRVIR1) imagery for the estimation of LAI in the boreal forest of Finland (estimation error of 0.28 m^2 m^{-2} vs. 0.39 m^2 m^{-2} respectively). Stankevich et al. (2017) also demonstrated a promising relationship (R^2 = 0.81) between the relative difference polarization index (RDPI) (derived from Sentinel-1 bands) and the LAI of temperate deciduous and mixed forest in Ukraine, which could not be observed in our study using VV, and VH polarized bands. The inferior result from SAR data in our study compared to other studies could be attributed to the differences in variables used (RDPI vs raw bands) and sensors (SPOT-HRVIR1 vs Sentinel-2) as well as the type of forest under consideration. More importantly, our models are species-specific, which might make them not directly comparable with the above studies.

The modeling of BAPH also demonstrated the superiority of optical variables (Sentinel-2) over the SAR variables. A similar observation for tree species BAPH modeling in Minnesota was reported by Wolter and Townsend (2011), where the performances of Radarsat (C-band SAR) and PALSAR (L-band SAR) were inferior to that of Landsat-5 data. In agreement with our findings, Shamsoddini (2012) reported the superior performance of SPOT-5 optical data for BAPH estimation as compared to

the Advanced Land Observation Satellite (ALOS) radar data for radiata pine (*Pinus radiata* L.) plantation in Australia. We specifically formulated Sentinel-2 multi-variable SVI, and individual spectral band models for both LAI and BAPH estimation to include the combined effect of different variables responsible for the determination of unique canopy traits. The aforementioned approach (multi-variable modeling) has already been practiced in several studies from the past for both LAI (Korhonen et al., 2017; Dube et al., 2019; Meyer et al., 2019) and BAPH (Wolter and Townsend, 2011; Astola et al., 2019; Ahmadi et al., 2020) estimation with a great promise, which can also be observed in our results as well (Table 2). The nRMSE was improved by 0.06 for RS-LAI and BF-LAI, using multi-variables in place of the single best predicting variable. Single variables were not tested for BAPH models since we used multi-temporal data for modeling.

The dominance of red-edge variables for modeling LAI (NDVIRE, IRECI, PSRI, and CRI1) and BAPH (S2REP, NDVIRE, IRECI, TCARI, PSRI, CRI1, and CRI2) in our study are in agreement with the findings from the previous investigations (Korhonen et al., 2017 Darvishzadeh et al., 2019). Majasalmi and Rautiainen (2016) found a moderate relationship between S2REP and/or IRECI and LAI, which does not align fully with our results (S2REP not selected). Nevertheless, NDVIRE (highly correlated with S2REP) was one of the most influential variables for the LAI models we developed. While the majority of the available studies on LAI estimation are focused on the broader spectrum of forest types (Majasalmi and Rautiainen, 2016; Korhonen et al., 2017; Meyer et al., 2019), our study proposes an algorithm that can estimate the LAI of specific tree species in mixed species forest conditions.

For BAPH estimation, the major role of red-edge variables in our study could be supported with several other studies with similar findings. Astola et al. (2019) compared the potential of Landsat-8 and Sentinel-2 variables for predicting several forest variables, including the BAPH in the boreal forest in southern Finland, where they concluded the advantage of red-edge bands in Sentinel-2, which resulted in better performance compared to Landsat-8. Furthermore, the study on mapping BAPH and other stand variables in northern Iran conducted by Ahmadi et al. (2020) also emphasizes the prime role of red-edge bands in estimating species-specific BAPH.

Site variables played a significant role in estimating both LAI and BAPH in our study. There have been very few investigations regarding the contribution of site variables to forest biophysical parameters estimation; however, their incorporation with remote sensing variables has produced improved results for similar studies (Bhattarai et al., 2020; Rahimzadeh-Bajgiran et al., 2020; Bhattarai et al., 2021). A recent study from Ahmadi et al. (2020) on estimating forest BAPH integrating site variables with Sentinel-2 data highlights the importance of elevation, which aligns with our findings. We found elevation to be one of the best variables for the estimation of LAI and relatively noteworthy in BAPH modeling (Fig. 3 and Table 2). In general, the inclusion of site variables in the final model (Sentinel-2 variables) improved the RMSE by 0.11–0.14 $m^2\ m^{-2}$ (nRMSE rose by 0.02–0.03) for LAI models, whereas by 0.02–0.09 $m^2\ ha^{-1}$ (nRMSE unchanged) for BAPH models. This finding generally reflects the diverse yet distinct factors that influence both stand species composition and productivity in this region (e.g. Rahimzadeh-Bajgiran et al., 2020; Bhattarai et al., 2021).

Comparing the performances of RF and MLP algorithms using the final sets of reduced variables (Fig. 3), RF model outperformed MLP in terms of the estimation error (Fig. 4); however, the RS-LAI model performed slightly better with MLP algorithm in terms of R^2. The results suggest that RF regression can model forest inventory attributes like LAI and BAPH comparatively better than ANN. Similar to our findings, several other researches have suggested using other algorithms like Gaussian process regression (Verrelst et al., 2015) and RF (Wei et al., 2017; Wang et al., 2018) over the ANN algorithm for the estimation of biophysical properties of vegetation. The better performance of RF in our study can likely be attributed to the robustness of the RF algorithm in handling high variability in the field data, which was evident in previous findings as well (Wang et al., 2018; Dube et al., 2019).

Both LAI and BAPH models were successfully cross-validated. Our R^2 values of 0.80–0.87 for BAPH models (Table 2) are relatively higher than some previous studies, while our RMSE values of 5.18–5.69 $m^2\ ha^{-1}$ are lower. For example, Townsend (2002) used a combination of radar images (ERS-1, JERS-1, and Radarsat) and Landsat imagery to achieve an overall validation R^2 of 0.59–0.87 and RMSE of 6.71–13.40 $m^2\ ha^{-1}$ for modeling BAPH of a mixed forest in North Carolina. Further, Wolter and Townsend (2011) used multi-sensor optical and SAR data to estimate the abundance of various coniferous and broadleaved species where the BAPH model for balsam fir attained an R^2 value of 0.78, which is comparatively lower than the R^2 value we achieved (0.87); nonetheless, the RMSE value they reported (2.26 $m^2\ ha^{-1}$; nRMSE unknown) is better than that of ours. Ahmadi et al. (2020) also achieved the highest R^2 of only 0.48 and RMSE of 8.12 $m^2\ ha^{-1}$ (inferior performance as compared to our model) for BAPH modeling using Bayesian additive regression trees technique.

The LAI models we developed attained a cross-validated R^2 of 0.75–0.78 and RMSE of 0.63–0.95 $m^2\ m^{-2}$. Our accuracy metrics were comparable and even better at times than other similar models reported in the literature (Korhonean et al., 2017; Dube et al., 2019; Meyer et al., 2019; Neinavaz et al., 2019; Qiao et al., 2019). Meyer et al. (2019) attained the best R^2 of 0.59 and RMSE of 0.88 $m^2\ m^{-2}$ for the LAI estimation of a temperate forest in Germany that is lower in terms of R^2, while relatively close to ours in terms of the RMSE. Furthermore, Korhonean et al. (2017) used Sentinel-2 and Landsat-8 to compare their potential to estimate LAI in the boreal forest of Finland with the highest cross-validated R^2 of 0.73, RMSE of 0.59, and nRMSE of 0.19, which is slightly better than our models in terms of RMSE but not in terms of R^2 (0.75–0.78) or nRMSE (0.12–0.16).

The empirical method we used for LAI and BAPH species-specific estimation using RF is promising for future reference; however, the comparisons we made between different studies acknowledges that they have used unique data collected from diverse sites and sensors. Given that there are limited studies on species-specific inventory parameter estimation using remote sensing techniques, our models can be useful for the estimation of species-specific LAI and BAPH in Maine as well as similar geographical regions. Nevertheless, we highly recommend that future studies evaluate the strength of other machine learning algorithms as well as SAR L-band and P-band data for LAI and BAPH modeling.

5. Conclusions

Sentinel-1 and −2 data are invaluable resources for remote sensing applications in forest resources, whose potential has been studied in this analysis for BAPH and LAI estimation for two economically and ecologically important conifers, RS and BF, in Maine, USA. Specifically, the diverse SVIs derived from Sentinel-2 imagery represent a multitude of canopy properties. Sentinel-1 SAR imagery was not found to be helpful in our study and needs additional evaluation in future studies. In particular, site variables were found promising in prediction models and should be considered more in future work.

Our study further demonstrated the superiority of RF over the MLP algorithm for estimating canopy biophysical parameters, implying the adequacy of the commonly used machine learning algorithms over deep learning algorithms for this particular application. However, the use of deep learning algorithms should still be considered for other datasets and variable combinations. The accuracy produced by our statistical models is encouraging. The models developed could be used to predict BF and RS abundance and understand SBW dynamics in light of the ongoing SBW outbreak in Canada and imminent arrival in Northeast, USA.

Funding

This research was funded by National Aeronautics and Space Administration (NASA) Cooperative Agreement awarded through the Maine Space Grant Consortium [Award No. 80NSSC19M0155] and the USDA National Institute of Food and Agriculture, McIntire-Stennis Project No. ME0-42119 through the Maine Agricultural and Forest Experiment Station (Maine Agricultural and Forest Experiment Station Publication Number 3887).

CRediT authorship contribution statement

Rajeev Bhattarai: Conceptualization, Formal analysis, Methodology, Validation, Writing – original draft, Writing – review & editing. **Parinaz Rahimzadeh-Bajgiran:** Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing. **Aaron Weiskittel:** Conceptualization, Resources, Writing – review & editing. **Saeid Homayouni:** Methodology, Writing – review & editing. **Tawanda W. Gara:** Writing – review & editing. **Ryan P. Hanavan:** Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Ahmadi, K., Kalantar, B., Saeidi, V., Harandi, E.K.G., Janizadeh, S., Ueda, N., 2020. Comparison of machine learning methods for mapping the stand characteristics of temperate forests Using Multi-Spectral Sentinel-2 Data. Remote Sens. 12 (18), 3019. https://doi.org/10.3390/rs12183019.

Astola, H., Häme, T., Sirro, L., Molinier, M., Kilpi, J., 2019. Comparison of Sentinel-2 and Landsat 8 imagery for forest variable prediction in boreal region. Remote Sens. Environ. 223, 257–273.

Bhattarai, R., Rahimzadeh-Bajgiran, P., Weiskittel, A., Meneghini, A., MacLean, D.A., 2021. Spruce budworm tree host species distribution and abundance mapping using multi-temporal Sentinel-1 and Sentinel-2 satellite imagery. ISPRS J. Photogramm. Remote Sens. 172, 28–40.

Bhattarai, R., Rahimzadeh-Bajgiran, P., Weiskittel, A., MacLean, D.A., 2020. Sentinel-2 based prediction of spruce budworm defoliation using red-edge spectral vegetation indices. Remote Sens. Lett. 11 (8), 777–786.

Breiman, L., 2001. Random forests. Machine Learning 45, 5–32.

Butler, B.J., 2018. Forests of Maine, 2017. Resource Update FS-160. Newtown Square, PA: US Department of Agriculture, Forest Service, Northern Research Station. 3p.

Chen, C., Rahimzadeh-Bajgiran, P., Weiskittel, A., 2021. Assessing spatial and temporal dynamics of a spruce budworm outbreak across the complex forested landscape of Maine, USA. Ann. Forest Sci. 78 (2), 1–14.

Cui, Z., Kerekes, J.P., 2018. Potential of red edge spectral bands in future Landsat satellites on agroecosystem canopy green leaf area index retrieval. Remote Sens. 10 (9), 1458. https://doi.org/10.3390/rs10091458.

Darvishzadeh, R., Wang, T., Skidmore, A., Vrieling, A., O'Connor, B., Gara, T., Ens, B., Paganini, M., 2019. Analysis of Sentinel-2 and RapidEye for retrieval of leaf area index in a saltmarsh using a Radiative Transfer Model. Remote Sens. 11 (6), 671. https://doi.org/10.3390/rs11060671.

Delegido, J., Verrelst, J., Alonso, L., Moreno, J., 2011. Evaluation of Sentinel-2 red-edge bands for empirical estimation of green LAI and chlorophyll content. Sensors 11 (7), 7063–7081.

DeRose, J.R., Seymour, R.S., 2010. Patterns of leaf area index during stand development in even-aged balsam fir – red spruce stands. Can. J. For. Res. 40, 629–637.

Dube, T., Pandit, S., Shoko, C., Ramoelo, A., Mazvimavi, D., Dalu, T., 2019. Numerical assessments of leaf area index in tropical savanna rangelands, South Africa using Landsat 8 OLI derived metrics and in-situ measurements. Remote Sens. 11 (7), 829. https://doi.org/10.3390/rs11070829.

Fang, H., Baret, F., Plummer, S., Schaepman-Strub, G., 2019. An overview of global leaf area index (LAI): methods, products, validation, and applications. Rev. Geophys. 57 (3), 739–799.

Fassnacht, K., Gower, S., Norman, J., Mcmurtric, R., 1994. A comparison of optical and direct methods for estimating foliage surface area index in forests. Agric. For. Meteorol. 71 (1-2), 183–207.

Foody, G.M., 2004. Supervised image classification by MLP and RBF neural networks with and without an exhaustively defined set of classes. Int. J. Remote Sens. 25 (15), 3091–3104.

Jain, A.K., Mao, J., Mohiuddin, K.M., 1996. Artificial neural networks: a tutorial. Computer 29, 31–44.

Korhonen, L., Hadi, Packalen, P., Rautiainen, M., 2017. Comparison of Sentinel-2 and Landsat-8 in the estimation of boreal forest canopy cover and leaf area index. Remote Sens. Environ. 195, 259–274.

Kuehne, C., Weiskittel, A.R., Wagner, R.G., Roth, B.E., 2016. Development and evaluation of individual tree-and stand-level approaches for predicting spruce-fir response to commercial thinning in Maine, USA. For. Ecol. Manage. 376, 84–95.

Majasalmi, T., Rautiainen, M., 2016. The potential of Sentinel-2 data for estimating biophysical variables in a boreal forest: a simulation study. Remote Sens. Lett. 7 (5), 427–436.

Mananze, S., Pôças, I., Cunha, M., 2018. Retrieval of maize leaf area index using hyperspectral and multispectral data. Remote Sens. 10 (12), 1942. https://doi.org/10.3390/rs10121942.

Manninen, T., Stenberg, P., Rautiainen, M., Voipio, P., Smolander, H., 2005. Leaf area index estimation of boreal forest using ENVISAT ASAR. IEEE Trans. Geosci. Remote Sens. 43 (11), 2627–2635.

Appiah Mensah, A., Petersson, H., Saarela, S., Goude, M., Holmström, E., 2020. Using heterogeneity indices to adjust basal area–leaf area index relationship in managed coniferous stands. For. Ecol. Manage. 458, 117699. https://doi.org/10.1016/j.foreco.2019.117699.

Meyer, L.H., Heurich, M., Beudert, B., Premier, J., Pflugmacher, D., 2019. Comparison of Landsat-8 and Sentinel-2 data for estimation of leaf area index in temperate forests. Remote Sens. 11 (10), 1160. https://doi.org/10.3390/rs11101160.

Neinavaz, E., Darvishzadeh, R., Skidmore, A., Abdullah, H., 2019. Integration of Landsat-8 thermal and visible-short wave infrared data for improving prediction accuracy of forest leaf area index. Remote Sens. 11 (4), 390. https://doi.org/10.3390/rs11040390.

O'Hara, K.L., Battles, J.J., 2020. Variation in leaf area index in complex mixed-conifer forests in California's Sierra Nevada: implications for stocking control. Forestry: An International Journal of Forest Research, 93, 641-651.

Qiao, K., Zhu, W., Xie, Z., Li, P., 2019. Estimating the seasonal dynamics of the leaf area index using piecewise LAI-VI relationships based on phenophases. Remote Sens. 11 (6), 689. https://doi.org/10.3390/rs11060689.

Rahimzadeh-Bajgiran, P., Hennigar, C., Weiskittel, A., Lamb, S., 2020. Forest potential productivity mapping by linking remote-sensing-derived metrics to site variables. Remote Sens. 12 (12), 2056. https://doi.org/10.3390/rs12122056.

Rahimzadeh-Bajgiran, P., Weiskittel, A., Kneeshaw, D., MacLean, D., 2018. Detection of annual spruce budworm defoliation and severity classification using Landsat imagery. Forests 9 (6), 357. https://doi.org/10.3390/f9060357.

Running, S.W., 1992. A bottom-up evolution of terrestrial ecosystem modeling theory, and ideas toward global vegetation modeling. In: Ojima, D. (Ed.), Modeling the Earth System. UCAR/Office for Interdisciplinary Earth Studies, Boulder, CO, pp. 263–280.

Shamsoddini, A., 2012. Radar backscatter and optical textural indices fusion for pine plantation structure mapping. ISPRS Ann. Photogr. Remote Sens. Spat. Inform. Sci. 7, 309–314.

Sprintsin, M., Karnieli, A., Berliner, P., Rotenberg, E., Yakir, D., Cohen, S., 2007. The effect of spatial resolution on the accuracy of leaf area index estimation for a forest planted in the desert transition zone. Remote Sens. Environ. 109 (4), 416–428.

Stankevich, S.A., Kozlova, A.A., Piestova, I.O., Lubskyi, M.S., 2017. Leaf area index estimation of forest using Sentinel-1 C-band SAR data. In 2017 IEEE Microwaves, Radar and Remote Sensing Symposium (MRRS), Kiev, Ukraine, 29-31 August 2017, 253-256.

Townsend, P.A., 2002. Estimating forest structure in wetlands using multitemporal SAR. Remote Sens. Environ. 79 (2-3), 288–304.

Verrelst, J., Rivera, J.P., Veroustraete, F., Muñoz-Marí, J., Clevers, J.G.P.W., Camps-Valls, G., Moreno, J., 2015. Experimental Sentinel-2 LAI estimation using parametric, non-parametric and physical retrieval methods–a comparison. ISPRS J. Photogramm. Remote Sens. 108, 260–272.

Wang, L.i., Chang, Q., Yang, J., Zhang, X., Li, F., Xu, Y., 2018. Estimation of paddy rice leaf area index using machine learning methods based on hyperspectral data from multi-year experiments. PLoS ONE 13 (12), e0207624.

Wei, C., Huang, J., Mansaray, L., Li, Z., Liu, W., Han, J., 2017. Estimation and mapping of winter oilseed rape LAI from high spatial resolution satellite data based on a hybrid method. Remote Sens. 9 (5), 488. https://doi.org/10.3390/rs9050488.

Weiskittel, A.R., Kershaw, J.A., Hofmeyer, P.V., Seymour, R.S., 2009. Species differences in total and vertical distribution of branch-and tree-level leaf area for the five primary conifer species in Maine, USA. For. Ecol. Manage. 258 (7), 1695–1703.

White, B., Ogilvie, J., Campbell, D.M.H.M.H., Hiltz, D., Gauthier, B., Chisholm, H.K.H., Wen, H.K., Murphy, P.N.C.N.C., Arp, P.A.A., 2012. Using the cartographic depth-to-water index to locate small streams and associated wet areas across landscapes. Canadian Water Resources Journal/Revue canadienne des ressources hydriques 37 (4), 333–347.

Wolter, P.T., Townsend, P.A., 2011. Multi-sensor data fusion for estimating forest species composition and abundance in northern Minnesota. Remote Sens. Environ. 115 (2), 671–691.

Wolter, P., Townsend, P., Sturtevant, B., Kingdon, C., 2008. Remote sensing of the distribution and abundance of host species for spruce budworm in Northern Minnesota and Ontario. Remote Sens. Environ. 112 (10), 3971–3982.

Yuan, J., Lv, X., Li, R., 2018. A speckle filtering method based on hypothesis testing for time-series SAR images. Remote Sens. 10 (9), 1383. https://doi.org/10.3390/rs10091383.